DOCUMENT RESUME

ED 245 593 HE 017 313

AUTHOR Warren, Jonathan R.
TITLE The Missing Evidence on Student Learning.
PUB DATE Mar 84
NOTE 11p.; Paper presented at the National Conference of the American Association for Higher Education (Chicago, IL, March 14-16, 1984).
PUB TYPE Viewpoints (120) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; *Aptitude Tests; *College Graduates; *College Programs; *Educational Quality; Higher Education; Information Sources; *Scores

ABSTRACT

The question of whether colleges are doing better or worse in educating their graduates compared with previous years is addressed. Attention is directed to recent trends in average test scores of students at graduation, as well as what the scores represent. Information is provided on changes since 1966 in several tests of academic aptitude taken by college graduates: the Graduate Record Examinations (GRE) General Test, the Law School Admission Test, and the Graduate Management Admission Test. Scholastic Aptitude Test scores from high school are also shown. Trends since 1968 are also indicated on eight subject tests. After briefly reviewing trends in test scores, it is noted that GRE scores provide information useful for graduate school admission, but they do not provide information on the educational quality of college programs. The following kinds of information that would provide acceptable evidence of educational quality are identified: compilation of the titles and catalog descriptions of courses undertaken by students who complete different programs; questions that make up the examinations given in the courses that constitute representative programs of study; and results of course examinations that indicate the scope and depth of student learning. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Jonathan R. Warren

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED245593

Session 87
Friday, March 16, 1984

THE MISSING EVIDENCE ON STUDENT LEARNING*

Jonathan R. Warren
Senior Research Scientist
Educational Testing Service

The quality of American higher education has been widely questioned in the last few years. Some of the phrases that have appeared in print are "quality in higher education is in jeopardy," it is "simply a morass," "shoddiness surrounds us," and less colorfully, "quality has noticeably declined in the past two decades." Yet when the basis for these and similar assertions is looked for, it's either hard to find or not very compelling.

Two kinds of information lie behind most criticisms of higher education's quality. One consists of standardized test scores, at entrance to college and then to graduate and professional schools. The other consists of faculty and administrator impressions--courses are not as rigorous as they used to be, students are not as committed, curricula have been watered down. In general, standards have dropped. Both kinds of evidence are weak, but I will limit myself to examining test scores.

There is no question that students in general are less well prepared than they used to be. They enter college weaker in verbal and quantitative abilities than five or ten or fifteen years ago, although that long-term decline seems to have bottomed out. Until recently, I believe that much of the decline was due to greater numbers of students scoring in the lower regions of the test score distribution, but the absolute number of high-scoring students has also dropped.

So entering students in general are not as capable as they used to be, remedial courses have indeed become swollen with poorly prepared students, and some courses that are not remedial have become less demanding. Yet before that can be accepted as evidence that educational quality at the undergraduate level has been eroded, or become "shoddy," a "morass," or even been placed in jeopardy, we would have to know what level of educational or intellectual performance students have reached by the time they leave college, whether at graduation or before. If they are leaving college at the same level of accomplishment as students of a decade or two ago, higher education will have cause for pride. Even if they are slightly less advanced than earlier students, they would represent a substantial educational accomplishment.

*Paper presented at Session 87 of the 1984 National Conference on Higher Education, sponsored by the American Association for Higher Education on Friday, March 16. Permission to quote restricted.

We do have test scores of students at graduation, though not at other points of exit. To find some meaning in them, we need two kinds of information. First, what have the trends been over recent years in average test scores, and second, what do those scores represent?

Figure 1 shows the changes since 1966 in several tests of academic aptitude taken by college graduates—The Graduate Record Examinations General Test, including Verbal and Quantitative scores, the Law School Admission Test, and the Graduate Management Admission Test. Also shown are the Scholastic Aptitude Test (SAT) scores of high school graduates four years earlier.

Whether colleges are doing better or worse in educating their graduates compared with previous years depends on which trend lines you look at and what consideration you give to the importance of preparation at entrance to college. The GRE Verbal scores have been in a continuous decline; the GRE Quantitative scores declined and then more than recovered; the GMAT scores declined and then leveled off; the LSAT scores have been in a continuous climb. The most reasonable interpretation, but probably not the whole story, is that the most capable college graduates have been shifting from graduate schools, which require the GRE, to law schools and, to a lesser extent, to graduate schools of business.

Even if the reasons for the different trends were known, however, they would not provide information about the quality of the educational programs the graduates had been through because the test graphed in Figure 1 don't measure what is taught in undergraduate programs. The GRE Verbal, LSAT, and GMAT scores all reflect reading comprehension, some form of reasoning ability, and in the case of the GMAT, quantitative ability that requires knowledge of high school math. None of these tests assesses anything that is an explicit objective of any college course, although four years of college experiences can be expected to increase a person's facility in verbal, reasoning, and quantitative activities to some extent simply through practice. The GMAT Bulletin of Information states explicitly, "The GMAT does not test specific knowledge obtained in college course work. It does not measure achievement in any particular subject area." Similarly, the GRE Quantitative scores reflect not college-level mathematical skills but high-school-level abilities.

The GRE Subject Tests are somewhat better although also quite limited. Figure 2 shows the trends since 1968 in scores on eight Subject Tests. Literature in English, History, and Political Science show declines that may have leveled off in the last two or three years. Psychology and Economics scores declined early in the period graphed and have been fairly stable for the past six or seven years. Scores in Mathematics, Biology, and Chemistry have climbed. If these test scores were to be interpreted as indicators of educational quality, we might infer that English, history, and political science departments were declining in quality while mathematics, chemistry, and biology departments were improving. But then if we take into account the declining level of preparation of entering students, as indicated by SAT scores, the

English, history, and political science departments seem to be just holding their own while the others are amazingly effective. The problem with both these interpretations is that even the GRE Subject Tests do not give information on which judgments of educational quality can be realistically based.

The Subject Tests are composed of questions on material expected to be taught in most departments across the country. If they were not, scores would depend to an unacceptable degree on the particular departments in which students had studied and the particular courses they had taken. The wide representativeness of the Subject Tests, however, keeps them from being able to reflect the particular strengths of any single department. And the upper-division courses that provide the depth expected in students' major fields are largely neglected for the same reason. They are too variable from college to college.

The GRE Board recognizes that limitation, as is indicated in its bulletin describing the History Test. "The problem of content coverage in a single history test is complex. It is almost impossible to delimit the field of history in area, in time, and in scope. Moreover, no common core of knowledge is required of all history majors in all colleges."[2] Because of that limitation, the History Test includes only U.S. and European history, concentrating on political and diplomatic history, U.S. history since about 1800, and European history after the 17th century. Neglected are history in the period of the Reformation or earlier, historical method, and Latin American, Asian, or Middle Eastern history. The patterns of study of many history majors and the curricula of many departments depart substantially from the patterns of content on the test. Similar comments could be made about each GRE Subject Test.

None of this invalidates the GRE tests for their intended purpose--the selection of applicants to a graduate program. Applicants who know the content of the Subject Tests have a reasonable framework of knowledge for graduate study in that field, and high scores on the General Test indicate facility for academic study. While GRE scores provide information useful in admission to graduate school, they do not and are not intended to provide information on educational quality.

We are left with no information about the quality of higher education or of individual institutions or departments other than unsupported impressions, such as those of college presidents, academic deans, or faculty members about the relative ranking of colleges or departments or their perceptions of whether educational quality has been improving, declining, or remaining stable. Better kinds of information are essential if we are to have any clear sense of the existing quality of higher education and the direction in which it is tending. Without that information, evaluations of improvement will be impossible and decisions about curriculum and instruction will be made on nothing more than intuition. Such decisions are probably often sound, but no one ever really knows.

Is better information available?

Several kinds of information can be developed that would provide acceptable evidence of educational quality, although no single kind will be adequate in itself and even in the aggregate they will never be complete. But those inevitable limitations should not deter us from using the information that can be provided.

Compilation of the titles and catalog descriptions of courses that appear on the transcripts of representative samples of graduates will provide one useful kind of information on what students who complete different programs can be expected to learn, with all the variability in the programs revealed. That information would be inaccurate and limited but informative in many ways. The extent to which historical method is emphasized in history, quantitative methods in political science, or physiology in psychology programs would become reasonably clear. The variety of patterns of content would also be clear. The differences, for example, between a program in materials engineering at Cal Tech and at an unselective state university would appear. While such differences would not in themselves reflect educational quality, since the two institutions can be presumed to have different purposes and serve different clienteles, but that kind of information is one start toward inferences about educational quality.

A step beyond the information in course titles and their catalog descriptions, which are not always up-to-date or accurate, is the information in the questions that make up the examinations given in the courses that constitute representative programs of study. Even without the examination results, the questions themselves would indicate much of what the faculty considered important and introduced into the courses. They would indicate what the students had encountered in their programs and, to a greater extent than is now possible, what they had learned.

Finally, the results of course examinations, if some care is taken in devising and administering the exams, can be aggregated into indicators of the scope and depth of student learning that would go far beyond anything now available. The problems in that approach to assessing educational quality are no different from those met in any such attempt. Differences in students' preparation and institutional purposes must be respected. The great variety of programs, even in a single field of study and across courses with identical titles must be accommodated. Fears of faculty that they will be unrealistically evaluated on the basis of students' efforts and abilities must be addressed. The examination questions themselves must be capable of being reliably graded. All four of these potential problems are manageable.

The knowledge or understanding tested by any particular examination question will be relevant to a large number of courses, often having different titles and even appearing in different departments. And it will usually not be relevant to every course carrying the same title. The solution is for faculty members to develop exam questions, or to

select those developed by others, that are pertinent to a particular course. Short essay questions requiring about ten minutes for a student response can be developed with sufficiently clear purposes and grading procedures that the collective results of a class of students will carry reliable information.

Faculty members who include two to four such questions on an exam consisting of eight to twelve will not be vulnerable to unfair judgments since they will control the content of the questions, and the questions will reflect only a small part of the material of the course. Yet the availability of results from other classes in other institutions to the same questions will give faculty members valuable information about how their students compare with others. The characteristics of the classes in which common examination questions are used--the selectivity of the institution, whether the class is an upper- or lower-division course, whether primarily for majors in the field or for general students, and the prerequisites usually completed by the students--will help faculty members put their own students' performance into a useful context. Finally, the aggregation of the results of such questions across a sample of courses in a department, an institution, or a group of institutions will provide badly needed information on educational quality.

Footnotes

1. American Council on Education. "Quality in American Higher Education." Unpublished background paper for the National Commission on Higher Education Issues. Washington, D.C.: American Council on Education, 1981.

Bonham, G. W. "Measuring Indicators of Quality." In E. Kuhos & S. V. Martorana (Eds.) Qualitative Methods for Institutional Research, New Directions for Institutional Research, No. 34. San Francisco: Jossey-Bass, 1982.

Enarson, H. L. "Quality—Indefinable but not Unattainable." Educational Record, 1986, 64(1), 7-9.

2. Educational testing Service. A Description of the History Test. Princeton, NJ.: Educational Testing Service, 1982.

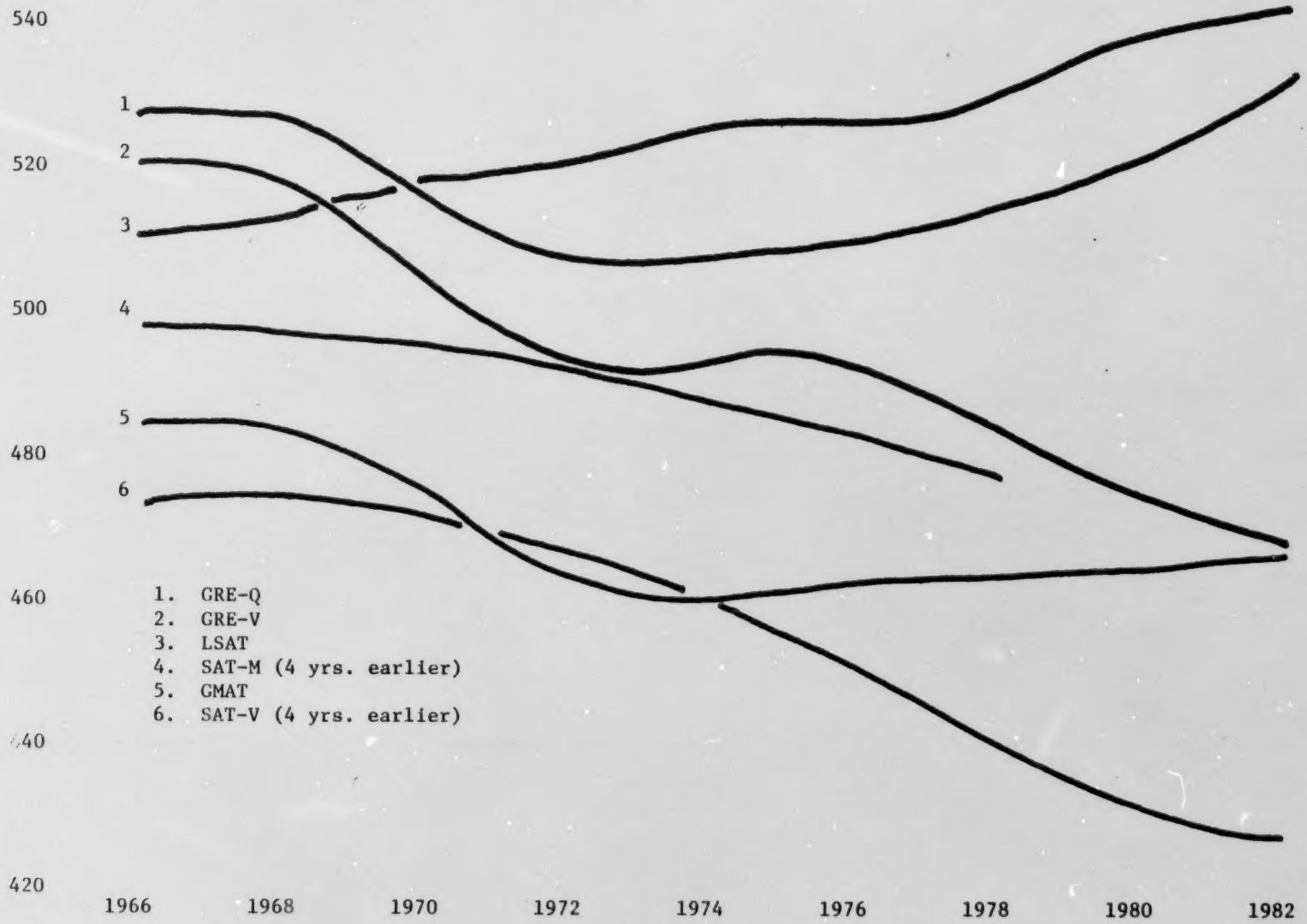

Figure 1. Trends in Test Scores

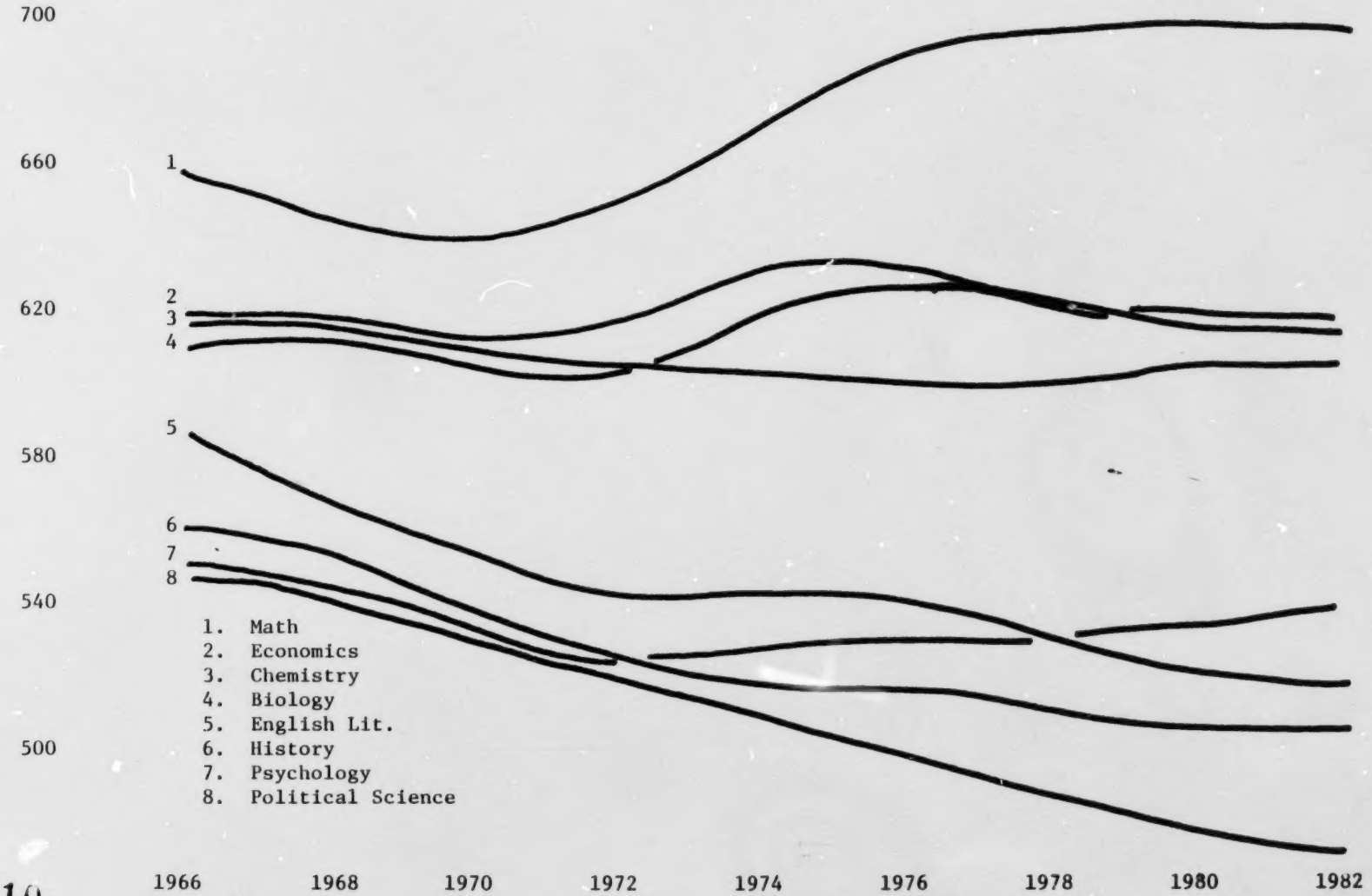

Figure 2. Trends in Achievement Test Scores